



University of Technology, Sydney

IN-DEPTH OPTIMIZATION OF STOCK MARKET DATA MINING TECHNOLOGIES

A Thesis Submitted for the Degree of
Doctor of Philosophy

By

Li LIN

Supervisor: Prof. Chengqi Zhang

Co-Supervisor: Dr. Shichao Zhang

Faculty of Information Technology
University of Technology, Sydney

AUSTRALIA

29 July 2006

Certificate

Author: **Li Lin**

Title: **In-Depth Optimization of Stock Market Data Mining Technologies**

Degree: Philosopher Doctor

Date: 29 July 2006

I certify that this thesis has not already been submitted for any degree and is not being submitted as part of candidature for any other degree.

I also certify that the thesis has been written by me and any help that I have received in preparing this thesis and all sources used have been acknowledged in this thesis.

Signature of Student

The image shows a handwritten signature in black ink. The signature consists of two parts: the first part is a stylized representation of the Chinese character '林' (Lin), and the second part is a stylized representation of the Chinese character '立' (Li). The signature is written in a cursive, flowing style.

Acknowledgements

I would like to extend my immense gratitude to my supervisor Professor Chengqi Zhang for his great selfless support and stimulating discussions, and for his prompting and careful reading of my draft which has been invaluable in the completion of this dissertation. He has provided an important mix of inspiration, sympathy and unfailing good humors, and he always gives me timely feedback for my research ideas, papers and dissertation drafts. I am lucky and happy to have been able to work with him during my PhD studies.

Also, I would like to thank the Faculty of Information Technology at University of Technology, Sydney (UTS) and CMCRC, SIRCA, AC3, SMARTS for providing me an excellent environment and the scholarship for my studies and research work. Especially for the CMCRC CEO Professor Mike Aitkin in the University of New South Wales, Australia, my co-supervisor Dr. Shichao Zhang and our team leader Dr. Longbing Cao in the University of Technology, Sydney, Australia.

Moreover, I appreciate all my other group members and my friends for plentiful help and happiness they have given to me, namely Ms. Li Liu, Mr. Jiaqi Wang, Mr. Wanli Chen, Mr. Jiarui Ni, Dr. Xiaowei Yan, Dr. Qingfeng Chen, Dr. Chunsheng Li. Mr. Zhenxing Qin, Mr. Yanchang Zhao, Ms. Dan Luo, A/Prof. Jie Lu, Dr. Guangquan Zhang, A/Prof. Maolin Huang, A/Prof. Sean He, A/Prof. Yuzhi Yan, A/Prof. Fenxi Zhang, Dr. Cynethia Nelson, Ms. Suneetha Uppu and more not listed here are also remembered in my heart.

Last but not the least, I am very grateful to my parents, my brothers, my wife Yinghui, my daughters Su and Si, who have been patient and supportive during my lives and studies in Australia. Their confidence and understanding are also a great encouragement to me in my academic endeavor.

I'd like to express my sincere thanks to those who once offered some help and concern to my study and life. My heart is filled with gratitude that words cannot express.

Contents

Certificate	i
Acknowledgements	ii
Contents	iv
Abstract	1
Chapter 1. Introduction	4
1.1 Problems	4
1.2 Solutions	8
1.3 Related Work	11
1.4 Significance	15
1.5 Structure of the Thesis	17
Chapter 2. Context	18
2.1 Data	18
2.2 Technical Trading Rules	20
2.2.1 Simple Moving Average	21
2.2.2 Filtered Moving Average	22
2.2.3 Enhanced Moving Average	23
2.2.4 Channel Break-Outs	25
2.2.5 Filter Rules	26
2.2.6 Support and Resistance	28
2.2.7 Other Trading Rules	29
2.3 Evaluation Criteria	29
2.3.1 Profit	30
2.3.2 Return	30
2.3.3 Sharpe Ratio	31
2.4 In-Sample and Out-of Sample Data Sets	32
2.5 Summary	33

Chapter 3. Domain Knowledge Integrated Applications	35
3.1 Problems	35
3.2 Domain-Driven Model	37
3.2.1 Related Concepts	37
3.2.2 Model	37
3.3 Human-Knowledge Interface	40
3.3.1 Domain Knowledge Database	40
3.3.2 Human-Machine Interface and Domain Knowledge In- tegration	41
3.4 Summary	44
Chapter 4. In-Depth Pattern Discovery and Related Applica- tions	45
4.1 Fundamental Concepts	45
4.2 DDID-PD Process Model	48
4.2.1 Constraint-Based Context	49
4.3 Mining In-Depth Patterns	50
4.3.1 Mining In-Depth Patterns	50
4.3.2 Human-Machine-Cooperated Interactive Knowledge Dis- covery	51
4.3.3 Loop-Closed Iterative Refinement	52
4.4 Mining Stock-Rule Pair in Real Stock Markets	53
4.4.1 Mining In-Depth Trading Rules (Sub-Domain)	53
4.4.2 Mining In-Depth Stock-Rule Pairs	57
4.5 Summary	58
Chapter 5. Optimized Algorithms	60
5.1 Standard Genetic Algorithm	60
5.1.1 Background	60
5.1.2 Solution	60
5.1.3 Conclusion	62
5.2 Robust Genetic Algorithms	64
5.2.1 Background	64
5.2.2 Robust Genetic Algorithms	64
5.2.3 Comparison	65
5.3 Fuzzy Set Methods	69
5.3.1 Fuzzy set	70
5.3.2 Output Literal Results	71
5.3.3 Evaluation and Conclusions	75
5.4 Multiple Criteria	76
5.4.1 Background	76

5.4.2	Solution	77
5.4.3	Conclusion	79
5.5	Summary	80
Chapter 6. Applications		82
6.1	Optimal Parameter Combination	82
6.2	Optimized Sub-Domain	83
6.3	Stock-Rules Pairs	86
6.3.1	Distribution of In-sample Set and Out-of-sample Set .	87
6.3.2	Support and Confidence	88
6.4	Relationship between Investment and the Number of Pairs . .	90
6.4.1	Profit, Return, Top Percentage Pairs and Investment .	90
6.4.2	Investment and the Number of Pairs	94
6.4.3	Stock-Rule Pair Return and Market Index Return . .	97
6.4.4	Maximal Return Point	101
6.4.5	Determination of Investment	103
6.5	Summary	106
Chapter 7. Evaluations		108
7.1	Financial Profitability	108
7.1.1	Economic Profitability	108
7.1.2	Sharpe Ratio	110
7.1.3	Predictability	112
7.2	Computational Performance	113
7.2.1	Execution Time	113
7.2.2	Memory	114
Chapter 8. Conclusions and Future Work		115
8.1	Conclusions	115
8.2	Future Work	117
Bibliography		119

Abstract

Stock trading is a number of stocks to be exchanged from one trader to another trader. It consists of a trader selling a number of stocks at a price and a volume, and another trader buying the same stocks at the same price and the same volume. Most traders want to buy a stock at a low price and to sell the stock at a high price in order to make a profit. However, it is difficult to know whether the current trading (buy/sell) price is low or high. Some researchers have presented technical trading rules which are mathematical formulas with many parameters to solve this problem, such as moving average rules, filter rules, support and resistance, channel break-out rules, and so on. All these rules are based on historical data to generate the best parameters and use the same parameters in future trading to make a profit. When the parameters of a trading rule are set properly, the trading rule can help the traders to make a profit (buy/sell at a low/high price). Experiments have shown that technical trading rules are profitable.

However, there are still some disadvantages and limitations to the technical trading rules in real stock market trading. First, the technical trading rules do not integrate domain knowledge (expert experiences and domain constraints, etc). For example, some trading rules pattern maybe only generate three signals during one year trading to get the most profit. However, the pattern is unreasonable and it is unprofitable in future trading, because the pattern is only a mathematical maximum, but it is impracticable in stock trading. Second, the output of a parameter for a trading rule is only one single value. Sometimes, it may be a noise so the trading rule is inapplicable in future trading. Third, present algorithms to calculate parameters of trading rules are inefficient. Most trades are performed through internet such that they can buy and sell stocks in online and a trade is completed in a second. Real markets are dynamic such that trading rules have to be updated all the time depending on changing situations (new data come

in, new parameters will be recomputed). Current enumerate algorithms waste too much time to get new parameters. However, a one-second short delay in real stock trading will lose the best trading chance. Fourth, when we evaluate the performance of a stock, we need not only to consider its performance (profit and return), but also to compare it to other stocks performance. At present, trading rules do not compare to the other stocks performance when they are selected to generate a signal, so the selected stocks or rules may be not the best ones. Fifth, in stock markets, there are many stocks and many trading rules. The problem is how to match and rank stocks and rules to combine a profitable and applicable pair. However, trading rules do not solve this problem. Lastly, trading rule techniques do not consider the sizes of investments. However, in real market trading, different investments will result in a different performance of a pair.

We propose in-depth data mining methodologies based on technical trading rules to overcome these disadvantages and limitations mentioned above. In this thesis, we present the solutions to combat the existing six problems.

To address the first problem, we designed a domain knowledge database to store domain knowledge (expert experience and domain constraints). During the computing procedure, we integrated domain knowledge and constraints. We observed the output more reasonable as we considered domain knowledge.

To address the second problem, we optimized a sub-domain output instead of a single value, in the sub-domain all combinations of parameter can get a near-best result. Moreover, in the sub-domain, some experienced traders can also set or micro-tune parameters by themselves and a better performance is guaranteed.

To address the third problem, we adopted genetic algorithms and robust genetic algorithms to improve the efficiency. Genetic algorithms and robust genetic algorithms can get a near-optimal result in an endurable execution time, and the result is also near to the best one.

To address the fourth problem, we applied fuzzy sets and multiple fitness functions to evaluate stocks. Because many factors influence the performance of a stock, it is necessary to create a multiple fitness function for genetic algorithms and robust genetic algorithms.

To address the fifth problem, we built a stock-rule performance table to rank stock-rule pairs and find the best matching pairs. The stock-rule pair results showed that the ranked performance is better than that of randomly matched pairs.

Finally, to address the sixth problem, we drew a graph of the relationship between investments and number of stock-rule pairs to search maximal points, and to decide the number of pairs for different sizes of investments.

In summary, the purpose of this thesis is to identify optimal methodologies in stock market trading, to make more profit with less risk for investors. The experimental results showed that the methodologies are more profitable and predictable.